Charles University, Faculty of Science
Department of Zoology
Study program: Zoology

Libor Mořkovský

Speciation genomics in nightingales

Genomika speciace u slavíků

Summary of the Doctoral thesis

Supervised by RNDr. Radka Reifová, Ph.D.
Prague 2019

Summary

Speciation is usually a slow process occurring over thousands to millions of years. This makes speciation research difficult because no direct observation or manipulation is possible. At best, we can gain some insight by inferring the population history and structure in very fine detail by investigating genetic markers in multiple individuals of the nascent species. Today, speciation research is in an unprecedented position thanks to the advent of high-throughput sequencing methods, which make it easier and cheaper than ever before to evaluate multiple markers in many individuals. Speciation is not a straightforward process that happens in the same way every time, but rather a phenomenon occurring when genetic and ecological circumstances acting in symphony ultimately lead to reproductive isolation of two subpopulations. This is why it is important to study multiple model systems to understand the general principles behind speciation. We worked with two species of nightingales (*Luscinia luscinia* and *L. megarhynchos*) that diverged approximately 1.8 Mya, likely due to glacial fluctuations in Europe. Our main goal was to use these new high-throughput sequencing methods to (1) detect interspecific hybrids between the species, (2) estimate levels of interspecific gene flow, (3) find areas of the nightingale genome that underlie reproductive isolation and, finally, (4) study the role of the Z chromosome in speciation. We designed a set of species-specific single nucleotide polymorphism markers, as well as primers for their detection, which can be used for hybrid identification. We found several areas of the genome with reduced interspecific gene flow, and the genes contained in those genome areas appear to be enriched in pathways related to oocyte development and in metabolic pathways. This is intriguing, especially because females are sterile in F_1 hybrids. We also found reduced levels of gene flow along the whole Z chromosome between the species, suggesting a large role of this chromosome in speciation. Our results suggest that this may be caused by higher levels of genetic drift on the Z chromosome due to stronger sexual selection acting on males. High levels of genetic drift on the Z chromosome may contribute to faster Z chromosome evolution and thus faster accumulation of hybrid incompatibilities on this chromosome.

Shrnutí

Speciace je obvykle pomalý proces, kde se čas měří v tisících až milionech let. To dělá výzkum speciace obtížným, protože "celou" speciaci nelze ani pozorovat, ani s ní experimentálně manipulovat. Nejvíce informací o speciaci nám poskytne podrobná populační struktura společně s rekonstruovanou populační historií vybraných populací, které považujeme za vznikající nebo nedávno vzniklé druhy. Díky nedávnému vzestupu nových sekvenačních metod lze nyní zjistit genotypy mnoha jedinců ve velkém rozlišení za přijatelnou cenu. Výzkum speciace tím získává nové obzory. Speciace totiž není jeden konkrétní proces, který se pokaždé odehraje podle stejného scénáře, ale spíše symfonie různých událostí, jež vedou v konečném důsledku k (trvalé) reprodukční izolaci dvou subpopulací. Proto je k pochopení celého fenoménu důležité zkoumat příklady u různých modelových systémů. My jsme si zvolili dva druhy slavíků (*Luscinia luscinia* a *L. megarhynchos*), které se oddělily před zhruba 1,8 milionem let, pravděpodobně díky posunům zalednění během střídání dob ledových. Naším hlavním cílem bylo pomocí nových sekvenačních metod (1) detekovat mezidruhové hybridy, (2) odhadnout úroveň genového toku mezi druhy, (3) najít oblasti v genomu slavíka, které podmiňují reprodukční izolaci mezi druhy a konečně (4) prozkoumat roli chromosomu Z při speciaci. Pro rozpoznání hybridů jsme navrhli sadu druhově specifických SNP (single nucleotide polymorphism) markerů společně s primery pro jejich detekci. V genomu slavíka jsme nalezli několik oblastí se sníženým genovým tokem mezi druhy. Geny v těchto oblastech jsou častěji součástí biochemických drah podílejících se na vývoji oocytu a metabolismu. To je zajímavé obzvláště proto, že samice jsou u F1 hybridů sterilní. Dále jsme vyhodnotili, že genový tok podél celého chromosomu Z je snížený, což naznačuje důležitou roli tohoto chromosomu ve speciaci. Naše výsledky naznačují, že to může být způsobeno silnějším genetickým driftem spojeným se silným pohlavním výběrem působícím na samce. Silný genetický drift může vést k rychlé evoluci chromosomu Z a tím pádem přispívat k rychlejšímu hromadění hybridních nekompatibilit na tomto chromosomu.

Introduction

Understanding how the Earth's biological diversity arose is one of the main aims of science. Ever since Alfred Russel Wallace and Charles Darwin conceived the idea that species emerge from varieties of their common ancestors (Wallace 1858; Darwin 1859), biologists have been trying to reconstruct the tree of life and to understand the mechanisms causing the bifurcations. To understand the mechanism of species origin, one has to first address the question: *What is a species?*

One intuitively knows what it means for an organism to belong to a particular species. The individual usually shares characteristic traits with other individuals of the same species and produces offspring which keeps these traits. This definition is, however, not universal, as there are many cryptic species, which cannot be distinguished from each other based on morphological traits. On the other hand there are examples when certain individuals of one species show very different traits from other individuals of the same species, and are in fact morphologically more similar to other species.

One of the most successful attempts at species definition in the past century was Ernst Mayr's **biological species concept** (BSC), famously formulated as "*Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups.*" (Mayr 1942). Mayr stresses the isolation aspect of speciation, going so far that the whole genome is expected to behave as a cohesive coadapted genetic unit. Thus species have to exist to protect these units because any exchange of any part of the genomes would destroy their integrity. The isolating mechanisms that are protecting the genome from mixing with others are at the center of speciation (under the prism of the BSC). Shortcomings of the BSC became apparent with analysis of recently produced molecular data where many well-established species (in the taxonomical view) seem to have little reproductive isolation.

Half a century later, James Mallet came up with a simplified view of the problem. He used Darwin's original definition where species are simply delineated by the absence of intermediate forms between two varieties, and incorporated new knowledge from genetics (Mallet 1995). He called it the **genotypic cluster definition**. The biggest advantage of this viewpoint is that species can be affected by gene flow, selection, and history independently, rather than being defined by these processes.

Since researchers have realized that reproductive isolation between many species is incomplete and gene flow can still occur between them, they started to find ways how to measure the levels of gene flow and how to find genomic regions where gene flow ceased. This knowledge can give us an insight into specific mechanisms of species origin, which is an ultimate goal of the speciation research. To address these questions, one has to first obtain genetic data from the incipient species ideally covering the whole genome. The main goal of this thesis was to generate such data for two nightingale species (*Luscinia megarhynchos* and *L. luscinia*) and use it to infer mechanisms and history of speciation in these two species.

Aims of the study

My aim was to use high-throughput sequencing methods to facilitate a genome wide ("genomics") view of mechanisms of speciation in two nightingale species. I had four specific goals:

1) Develop methods to detect interspecific hybrids between the two species.
2) Estimate locus-specific levels of interspecific gene flow.
3) Find areas of the nightingale genome that underlie reproductive isolation.
4) Study the mechanisms behind the large Z effect in speciation of nightingales.

Material and methods

We studied hybridization and speciation in two sister species of passerine birds, the Common Nightingale (*Luscinia megarhynchos*) and the Thrush Nightingale (*L. luscinia*). These two species diverged 1.8 Mya (Storchová et al. 2010) and currently co-occur in a zone of secondary contact running across Central and Eastern Europe (**Figure 1**). The species are very similar morphologically and ecologically (Reif et al. 2015), although partial habitat segregation associated with divergence in bill size has occurred since secondary contact, in response to interspecific competition in sympatric populations (Reif et al. 2018; Sottas et al. 2018).

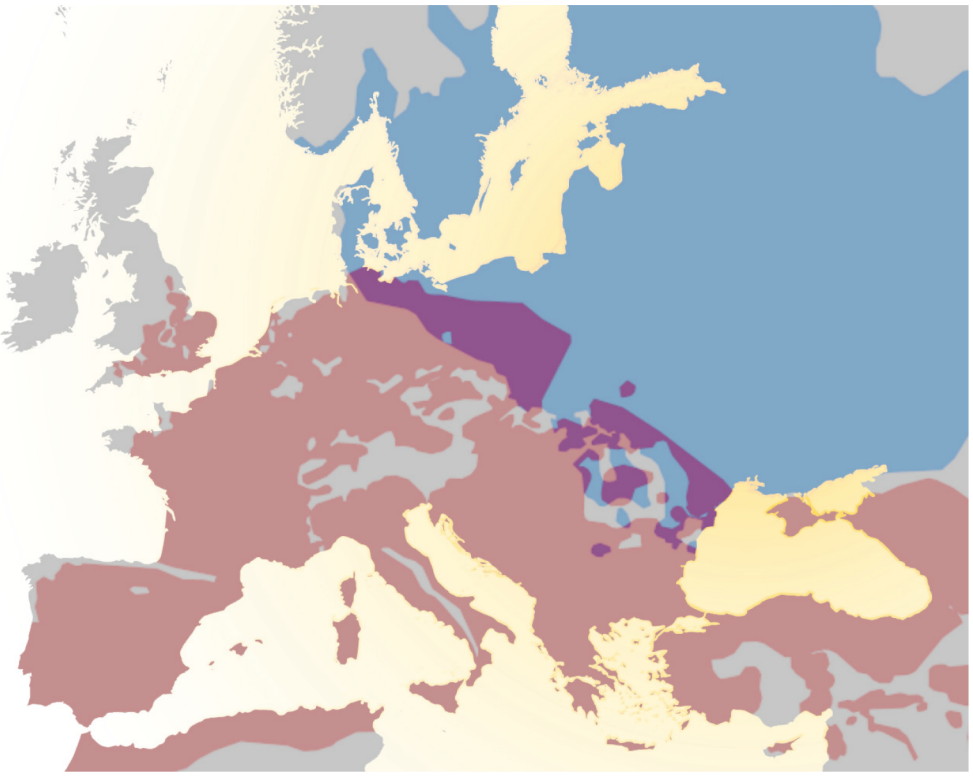

Figure 1 | **Breeding distribution** of **(red)** Common Nightingale (*Luscinia megarhynchos*) and **(blue)** Thrush Nightingale (*L. luscinia*) in Europe (overlap in violet). Data for species' geographic ranges were taken from BirdLife International (2019).

The two species sing a different song in allopatric areas but in sympatry, *L. luscinia* males very often incorporate song types from *L. megarhynchos* into their repertoires (Sorjonen 1986; Lille 1988). Recently, it was also found that nightingale species differ regarding sperm length, with *L. megarhynchos* having longer sperm when compared to *L. luscinia* (Albrecht et al. 2019), indicating that some form of gametic isolation may be present in these species. Despite the relatively strong assortative mating, the species occasionally hybridize in sympatry and produce viable hybrid progeny. Following Haldane's Rule, F_1 hybrid females are sterile, but F_1 hybrid males are fertile (Stadie 1991; Reifová et al. 2011). Previous work using coalescent models of isolation with migration showed that gene flow has occurred between the species and that gene flow has been significantly higher on the autosomes than on the Z chromosome, consistent with a "large Z effect" (Storchová et al. 2010).

Methods

We decided to start with sequencing the transcriptomes of multiple individuals collected in allopatry. After initial analysis, we used this data as a starting point for several investigations regarding the details of the speciation process.

To address goal **(1)**, we determined a panel of species-specific single nucleotide polymorphisms (SNPs) and designed PCR and genotyping primers for their detection. To detect different the hybrid categories we used the program *NewHybrids* (Anderson and Thompson 2002).

To address goal **(2)**, we pre-selected introns by comparing zebra finch and chicken genomes (ICGSC 2004; Warren et al. 2010). From the alignment we estimated intron length and availability of conserved regions in the flanking exons. We designed primers for sequencing of the selected introns and then further analyzed the intronic sequences in the framework of isolation-with-migration model (Hey and Nielsen 2004).

We used the whole transcriptome data set to find genomic regions with reduced gene flow via a genome scan—using zebra finch genome as a reference (Warren et al. 2010)—and analyzed their gene content. We analyzed functional categories of genes in islands using the KEGG Pathways gene lists (Kanehisa and Goto 2000) to address goal **(3)**.

Finally, we used the whole transcriptome data set to analyze patterns of molecular evolution on autosomes and the Z chromosome using indicators of population differentiation—for example F_{ST} and D_{XY} (Charlesworth 1998)—and selection—the MK test (McDonald and Kreitman 1991) and NI (Rand and Kann 1996) to address goal **(4)**.

Results and discussion

My research goals were:

1) Develop methods to detect interspecific hybrids between the two species.
2) Estimate locus-specific levels of interspecific gene flow.
3) Find areas of the nightingale genome that underlie reproductive isolation.
4) Study the mechanisms behind the large Z effect in speciation of nightingales.

The need to detect interspecific hybrids (**goal 1**) arose when we were investigating the influence of genotype on behavior by studying bird song, which is expected to play an important role in mate choice. To do so, we had to identify hybrids in the sympatric population. I developed a bioinformatics pipeline that uses transcriptome data to find the most suitable species-specific SNPs, and then creates a set of PCR and genotyping primers that can be used to detect these SNPs in a large number of individuals (Mořkovský et al. 2015). The selected SNP markers were then used to detect F_1 and backcross hybrids using the program NewHybrids in a later study of ours (Vokurková et al. 2013). We found that hybridization and introgression do not have any effect on observed changes in song patterns in sympatric populations.

To detect locus-specific levels of historical gene flow on the Z chromosome (**goal 2**), we fitted an isolation-with-migration model via the IM program to multiple intronic loci evenly spaced along the chromosome. To obtain the intronic sequences, I created a tool that aligned chicken and zebra finch orthologous exons and suggested primer locations in the most conserved regions. We found that there was only limited detectable gene flow on the Z chromosome. The gene flow was asymmetrical, mostly in the direction from *L. luscinia* to *L. megarhynchos* (Janoušek et al. 2019), which could mean that the Z chromosome in *L. megarhynchos* accumulated more hybrid incompatibilities. Interestingly, using transcriptomic data from both species, we found increased levels of genetic drift on the Z chromosome of *L. megarhynchos*, very likely caused by more intense postcopulatory selection acting on the males of this species. Our results suggest that lower effective population size of the Z chromosome and increased levels of genetic drift could contribute to the large effect of this chromosome in speciation (**goal 4**).

To determine if the nightingale genome exhibits the "islands of divergence" pattern, we analyzed the transcriptomes of 16 individuals (**goal 3**). Building on the transcriptome analysis pipeline (Mořkovský et al. 2015), I created contig scaffolds based on mapping of the contigs to the zebra finch genome. I then assessed divergence between the species in sliding windows, using both relative (F_{ST}) and absolute (D_{XY}) measures. We found 67 regions with elevated divergence. Enrichment analysis of the genes in islands suggested a possible association with oocyte development KEGG pathways, which is an interesting result given the sterility of hybrid females. Another

enriched KEGG pathway was the "metabolic pathways" collection. This could be explained by the different breeding distributions of the two nightingale species (*L. luscinia* breeds in a more northerly region than does *L. megarhynchos*), which could in principle lead to differences in metabolic rates between the two species (Mořkovský et al. 2018).

Conclusions

Together, the papers presented here advance our understanding of the nightingale model system in several areas that are currently deemed important for understanding speciation: the role of **sex chromosomes** in reproductive isolation, the existence and gene content of "**islands of divergence**", and the relationship between **genotype and behavior** leading to reproductive isolation. Similar facets of speciation are being actively investigated in other model systems, which means that we have contributed to a growing body of directly comparable evidence that is facilitating a deeper understanding of the speciation phenomenon.

References

Albrecht T, Opletalová K, Reif J, Janoušek V, Piálek L, Cramer ERA, Johnsen A, Reifová R. **2019**. Sperm divergence in a passerine contact zone: Indication of reinforcement at the gametic level. *Evolution.* **73**(2):202–213. doi:10.1111/evo.13677.

Anderson EC, Thompson EA. **2002**. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics.* **160**(3):1217–1229.

BirdLife International. **2019**. IUCN Red List for birds.

Charlesworth B. **1998**. Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol.* **15**(5):538–543. doi:10.1093/oxfordjournals.molbev.a025953.

Darwin C. **1859**. On the origin of species by means of natural selection. *John Murray, London.*

Hey J, Nielsen R. **2004**. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics.* **167**:747–760. doi:10.1534/genetics.103.024182.

ICGSC (International Chicken Genome Sequencing Consortium). **2004**. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature.* **432**(7018):695. doi:10.1038/nature03154.

Janoušek V, Fischerová J, Mořkovský L, Reif J, Antczak M, Albrecht T, Reifová R. **2019**. Postcopulatory sexual selection reduces Z-linked genetic variation and might contribute to the large Z effect in passerine birds. *Heredity.* **122**(5):622. doi:10.1038/s41437-018-0161-3.

Kanehisa M, Goto S. **2000**. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1):27–30. doi:10.1093/nar/28.1.27.

Lille R. **1988**. Species-specific song and mixed singing of Nightingale and Thrush Nightingale (*Luscinia megarhynchos*, *L. luscinia*). *J Ornithol.* **129**(2):133–159.

McDonald JH, Kreitman M. **1991**. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature.* **351**(6328):652. doi:10.1038/351652a0.

Mořkovský L, Janoušek V, Reif J, Řídl J, Pačes J, Choleva L, Janko K, Nachman MW, Reifová R. **2018**. Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. *Mol Ecol.* **27**(4):949–958. doi:10.1111/mec.14479.

Mořkovský L, Pačes J, Řídl J, Reifová R. **2015**. Scrimer: designing primers from transcriptome data. *Mol Ecol Resour.* **15**(6):1415–1420. doi:10.1111/1755-0998.12403.

Rand DM, Kann LM. **1996**. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol.* **13**(6):735–748. doi:10.1093/oxfordjournals.molbev.a025634.

Reif J, Jiran M, Reifová R, Vokurková J, Dolata PT, Petrusek A, Petrusková T. **2015**. Interspecific territoriality in two songbird species: potential role of song convergence in male aggressive interactions. *Anim Behav.* **104**:131–136. doi:10.1016/j.anbehav.2015.03.016.

Reif J, Reifová R, Skoracka A, Kuczyński L. **2018**. Competition-driven niche segregation on a landscape scale: Evidence for escaping from syntopy towards allotopy in two coexisting sibling passerine species. *J Anim Ecol.* **87**(3):774–789. doi:10.1111/1365-2656.12808.

Reifová R, Kverek P, Reif J. **2011**. The first record of a female hybrid between the Common Nightingale (Luscinia megarhynchos) and the Thrush Nightingale (*Luscinia luscinia*) in nature. *J Ornithol.* **152**(4):1063–1068. doi:10.1007/s10336-011-0700-7.

Sanger F, Nicklen S, Coulson AR. **1977**. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci.* **74**(12):5463–5467.

Sorjonen J. **1986**. Mixed Singing and Interspecific Territoriality – Consequences of Secondary Contact of Two Ecologically and Morphologically Similar Nightingale Species in Europe. *Ornis Scand Scand J Ornithol.* **17**(1):53–67. doi:10.2307/3676753.

Sottas C, Reif J, Kuczyński L, Reifová R. **2018**. Interspecific competition promotes habitat and morphological divergence in a secondary contact zone between two hybridizing songbirds. *J Evol Biol.* **31**(6):914–923. doi:10.1111/jeb.13275.

Stadie C. **1991**. Erdsanger I; Nachtigall und Sprosser. *Eur Vogelwelt.* **3**:130–189.

Storchová R, Reif J, Nachman MW. **2010**. Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evol Int J Org Evol.* **64**(2):456–471. doi:10.1111/j.1558-5646.2009.00841.x.

Vokurková J, Petrusková T, Reifová R, Kozman A, Mořkovský L, Kipper S, Weiss M, Reif J, Dolata PT, Petrusek A. **2013**. The Causes and Evolutionary Consequences of Mixed Singing in Two Hybridizing Songbird Species (*Luscinia* spp.). *PLoS ONE.* **8**(4): e60172. doi:10.1371/journal.pone.0060172.

Wallace AR. **1858**. On the tendency of varieties to depart indefinitely from the original type. *In: Proceedings of the Linnean Society of London.* **3**:53–62.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, et al. **2010**. The genome of a songbird. *Nature.* **464**(7289):757–762. doi:10.1038/nature08819.

Curriculum vitae

Mgr. Libor Mořkovský (*1985)

Education

since 2010 — Ph.D. program in Zoology
Faculty of Science, Charles University in Prague
Doctoral thesis: Speciation genomics in nightingales.

2008–2010 — Master's degree in Developmental biology
Faculty of Science, Charles University in Prague
Master's thesis: Characterization of Z chromosome gene content in birds.

2005–2008 — Bachelor's degree in Molecular biology and biochemistry
Faculty of Science, Charles University in Prague
Thesis: Evolution of sex chromosome gene content.

Biology related work experience

2016–2018 — Researcher and teacher
Department of Zoology,
Faculty of Science, Charles University in Prague

2012–2015 — Bioinformatician
Institute of Vertebrate Biology, Czech Academy of Sciences

2011–2013 — Researcher
Department of Zoology,
Faculty of Science, Charles University in Prague

Teaching

2015–2019 — Genomes and Unix (bioinformatics course)
Charles University in Prague

2016 — Software for Scientists: R course (lecturer)
Institute of Molecular Genetics, Czech Academy of Sciences

2006–2014 — Member of the Biology Olympiad authoring team

2007–2008 — Manager of the Biology Olympiad authoring team

Supervision of students

2014–2016 — Vojtěch Tuma
Department of Theoretical Computer Science and Mathematical Logic, Faculty of Mathematics and Physics, Charles University
Diploma thesis: Sumarizace genových expresních čipů z volně žijících druhů.

Grant projects

2012–2015 Charles University Grant Agency project No. 632712,
Project title: The role of interspecific interactions in the evolution of reproductive isolation in two species of nightingales.

List of publications

The thesis is based on the following publications:

Mořkovský L, Pačes J, Rídl J, Reifová R. **2015**. Scrimer: Designing primers from transcriptome data. *Mol Ecol Resour.* **15**(6):1415–1420. doi:10.1111/1755-0998.12403.

Journal Impact Factor (2015): 5.298

Mořkovský L, Janoušek V, Reif J, Rídl J, Pačes J, Choleva L, Janko K, Nachman MW, Reifová R. **2018**. Genomic islands of differentiation in two songbird species reveal candidate genes for hybrid female sterility. *Molecular Ecology.* **27**(4):949–958. doi:10.1111/mec.14479.

Journal Impact Factor (2017): 6.131

Janoušek V, Fischerová J, **Mořkovský L**, Reif J, Antczak M, Albrecht T, Reifová R. **2019**. Postcopulatory sexual selection reduces Z-linked genetic variation and might contribute to the large Z effect in passerine birds. *Heredity.* **122**(5):622. doi:10.1038/s41437-018-0161-3.

Journal Impact Factor (2017): 3.872

Vokurková J, Petrusková T, Reifová R, Kozman A, **Mořkovský L**, Kipper S, Weiss M, Reif J, Dolata PT, Petrusek A. **2013**. The causes and evolutionary consequences of mixed singing in two hybridizing songbird species (*Luscinia* spp.). *PLoS ONE.* **8**(4):e60172. doi:10.1371/journal.pone.0060172.

Journal Impact Factor (2013): 3.534

Other publications:

Matzke-Karasz R, Neil JV, Smith RJ, Symonová R, **Mořkovský L**, Archer M, Hand SJ, Cloetens P, Tafforeau P. **2014**. Subcellular preservation in giant ostracod sperm from an early Miocene cave deposit in Australia. *Proc R Soc B.* **281**(1786):20140394. doi:10.1098/rspb.2014.0394.

Journal Impact Factor (2014): 5.051

Symonová R, Majtánová Z, Arias-Rodriguez L, **Mořkovský L**, Kořínková T, Cavin L, Pokorná MJ, Doležálková M, Flajšhans M, Normandeau E, et al. **2017**. Genome Compositional Organization in Gars Shows More Similarities to Mammals than to Other Ray-Finned Fish. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution.* **328**(7):607–619. doi:10.1002/jez.b.22719.

Journal Impact Factor (2017): 2.432